Dual Low-Rank Decompositions for Robust Cross-View Learning

Zhengming Ding, Yun Fu, *Senior Member, IEEE*

***Abstract*—Cross-view data are very popular contemporarily, as different view-points or sensors attempt to richly represent data in various views. However, cross-view data from different views present a significant divergence, that is, cross-view data from the same category have a lower similarity than those in different categories but within the same view. Considering that each cross-view sample is drawn from two intertwined manifold structures, i.e., class manifold and view manifold, in this paper, we propose a Robust Cross-View Learning framework (RCVL) to seek a robust view-invariant low-dimensional space. Specifically, we develop a dual low-rank decomposition technique to unweave those intertwined manifold structures from one another in the learned space. Moreover, we design two discriminative graphs to constrain the dual low-rank decompositions by fully exploring the prior knowledge. Thus, our proposed algorithm is able to capture more within-class knowledge and mitigate the view divergence to obtain a more effective view-invariant feature extractor. Furthermore, our proposed method is very flexible in addressing such a challenging cross-view learning scenario that we only obtain the view information of the training data while with the view information of the evaluation data unknown. Experiments on face and object benchmarks demonstrate the effective performance of our designed model over the state-of-the-art algorithms.**

***Index Terms*—Cross-view Learning, Low-rank Modeling; Graph Embedding**

I. Introduction

CROSS-VIEW learning has caught an increasing attention during the past decades [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], as cross-view data are frequently observed around the world when data are collected from various view-points [2], [15], [16] or different sensors [1], [6], [7]. Although different views could facilitate better data representations, it leads to the difficulty that the same class data attempt to be lying in various distributions. Take cross-pose face recognition as an example, the pose variations are in 3D space, however, the image captures only 2D appearances. When the face pose changes, some visible parts may even be self-occluded, while some invisible parts may appear. It leads to the special phenomenon that the similarity between two different persons with similar pose is higher than the similarity between the same person across different poses (Figure 1), where we adopt the pre-trained deep structure with center loss [17] to extract the features and calculate the cosine similarity. However, we notice that large view variance could not still be solvable.

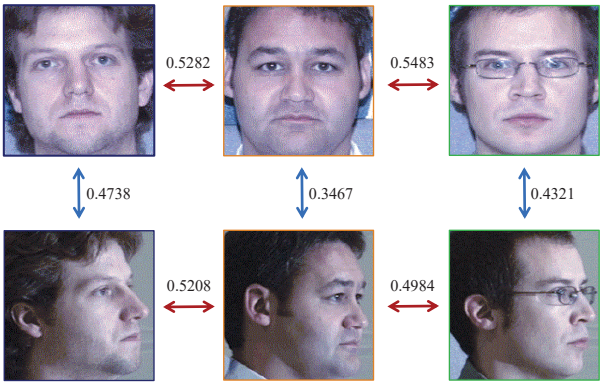


Fig. 1. Illustration of cross-view face, in which the first row shows the frontal faces of three persons while the second row represents the profile faces of the same three persons from CMU-PIE face dataset. Here we calculate the cosine similarity of each pair with 1024-dim deep features [17].

Recent research efforts on robust feature learning, e.g., sparse representation [18], [19] and low-rank representation [20], [21], manage to model the view variances as the noise parts so that they recover the clean data by detecting and removing noises. Specifically, sparse representation is robust to corrupted data, and has brought in impressive performance for face recognition [18], [19]. However, sparse representation ignores the underlying global structure within the data. Later on, low-rank representation has received great interest, which is able to capture the intrinsic structure within the data [20], [21]. Following this, some works even integrated low-rank and sparse representation as a whole to enhance the learning tasks [22], [23]. Most recently, low-rank constraint [21] has been well exploited in robust subspace learning [24], [25], [26], [13], which effectively incorporates dimensionality reduction and data structure recovery into a joint framework by leveraging the merits of both. However, the view variances would be much larger, thus, they cannot be treated as sparse noise and removed out with recent sparse or low-rank modeling algorithms [18], [20], [21].

Interestingly, we observe that there exist two different manifold structures within the cross-view data intertwined in the high-dimensional space. Specifically, one sample lies in two manifold structures, for example, the frontal face in Fig. 1 would belong to its own class manifold but it also lies in frontal pose manifold. However, the pose variance would hinder the classification task. Thus, it is essential to decompose such two structures by mitigating the view variance for cross-view learning. Moreover, most recent works [27], [28], [29],

Zhengming Ding is with the Department of Computer, Information and Technology, Indiana University-Purdue University Indianapolis, 420 University Blvd Indianapolis, IN 46202, USA. E-mail: zd2@iu.edu

Yun Fu is with the Department of Electrical and Computer Engineering and the College of Computer and Information Science, Northeastern University, Boston, MA, 02115 USA. E-mail: yunfu@ece.neu.edu.

This is the author's manuscript of the article published in final edited form as:
Ding, Z., & Fu, Y. (2018). Dual low-rank decompositions for robust cross-view learning. IEEE Transactions on Image Processing, 28(1), 194-204.
https://doi.org/10.1109/TIP.2018.2865885

[30] explore dual low-rank decomposition techniques to fight off heavy corruption difficulty, in which they also constrain the noisy part to be low-rank. However, they do not consider the view structure as an individual low-rank structure in cross-view learning. They both assume the data are from one single subspace following the idea of RPCA [20]. Furthermore, although they attempt to seek a robust subspace, they only learn a rotation matrix and cannot make merit by building an effective low-dimensional space.

In this paper, we present a novel Robust Cross-View Learning (RCVL) algorithm to build a robust view-invariant feature extractor via a dual low-rank decomposition technique. Since there exist two intertwined structures lying in cross-view data, it is the key to capture more intra-class knowledge while mitigating the view divergence within the same class. To our best knowledge, it is the first work to exploit dual low-rank decomposition strategy for cross-view learning. To this end, we highlight our main contributions in four folds as follows:

- Dual low-rank decomposition technique is designed to unweave two intertwined manifolds underlying the cross-view data. Hence, a more effective view-invariant space is built to preserve more discriminative information for recognition task. In this way, the view variance would be well addressed during the robust subspace learning.
- Two discriminative graphs are developed to supervise the dual low-rank decompositions with label and view information. This practice would alleviate the classification task by preserving more discriminative information while reducing view-variance impact within class.
- We adopt a novel rank approximation term to address the rank minimization problem so that our algorithm could achieve a much closer rank to its real value, by comparing with nuclear norm.
- Our proposed algorithm is a more flexible cross-view learning method, which can be easily generalized to the challenge in which the view information of test data is unavailable. In this scenario, traditional multi-view learning approaches [3], [5], [6], [31], [32], [33] would be invalid, since they only learn multiple view-specific transformations.

The remaining sections of this paper are organized as follows. In Section II, we provide a brief review of the related works and highlight the differences. We present our novel dual low-rank subspace algorithm in Section III, as well as the solution and complexity analysis of our method. Experimental analyses are provided in Section IV, followed with the conclusion in Section V.

II. Related Work

In this section, we first briefly revisit cross-view learning, then we highlight the differences between those related works and ours.

Cross-view learning aims to solve the problems when we have the data from different views, e.g., view-points, sensors, or feature types. The popular topics belonging to this scenario include cross-pose image classification [2], [15], heterogeneous image classification [34], and domain adaptation [7], [35], [36]. There is a recent survey discussing multi-view learning and domain adaptation in terms of different data organization, problem settings and research goals [12]. In general, two categories of techniques are explored to analyze the cross-view data, e.g., feature learning [5], [1], [6], [7], [24], [37] and classifiers adaptation [38], [39]. Our designed method lies in the feature learning category (i.e., subspace learning).

Traditional cross-view subspace approaches [2], [6], [31] were designed to learn multiple view-specific projections, which transform various views into a shared view-invariant space. Following this, the most representative one is Canonical Correlation Analysis (CCA), which learned two coupled projections to align two-view data into a shared low-dimensional space [40]. When facing more than two views, multi-view CCA [32] was proposed by extending CCA to multiple view cases. Following this, Zhao et al. proposed a deep non-negative matrix factorization model for multi-view data analysis by seeking multiple deep neural networks [33]. However, the key drawback for those algorithms is that they mainly deal with the multi-view learning problems by assuming view information for evaluation data is known. Thus, they would be invalid when we have no access to the view information of test data ahead of time.

Most recently, CNN-based deep learning approaches attract a lot of interest in view-invariant feature learning [41], [42]. However, all these deep methods need a huge number of labeled data to train the deep architecture, meanwhile they assume that deep structure is invariant to different domains and transferable to various tasks. However, the latest researches reveal that feature adaptability drops extremely in top layers with view divergence enlarged [43]. That is to say, deep structure cannot perfectly handle the large view variance within cross-view data (as we can see the similarity results from Figure 1).

Our work manages to seek a robust view-invariant subspace to well deal with the view divergence. Specifically, we develop a dual low-rank decomposition framework by assuming cross-view data are lying in two intertwined structures in term of the original space. Through dual low-rank decompositions, the intertwined structures would be well disentangled so that we could preserve more within-class knowledge while removing view variance influence. This paper is the extension of our previous conference work [26]. Differently, we adopt a novel rank approximation term to replace nuclear norm to fight off the rank minimization problem, so that we could achieve much closer rank to the real rank of data. Moreover, we explore a different solution to the subspace learning by directly using Eigen-decomposition instead of two-step optimization, which helps obtain optimal solution more efficiently. In addition, we conduct more experiments to testify our approach, e.g., evaluation on deep features and large-scale dataset, the effectiveness of dual low-rank decompositions and robustness to different-level noise.

III. The Proposes Algorithm

In this section, we first list our novel Robust Cross-View Learning through dual low-rank decompositions. Then, we provide an efficient solution as well as complexity analysis.

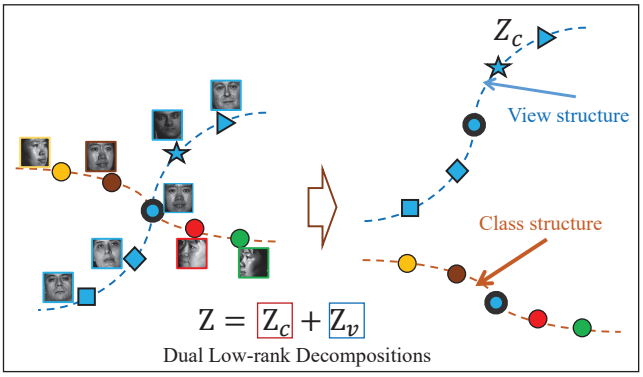


Fig. 2. Illustration of our dual low-rank decompositions. There exist two manifold structures intertwined, one for class information and the other for view information. Note that the same shape means the points are from the same class, whilst the same color denotes the points are from the same pose. Here we show 5 poses and 5 classes. Z is decomposed into two low-rank parts Z_c and Z_v for class structure and view structure, respectively.

A. Dual Low-rank Decompositions

Assume there is a set of cross-view data $\{X, y\} = \{(X_1, y_1), \cdots, (X_k, y_k)\}$ from k views. While each view $X_i \in \mathbb{R}^{d\times n_i}$ includes c categories, where d is the feature dimensionality and n_i is the sample size per view ($n = \sum_{i=1}^{k} n_i$). Conventional low-rank modelings [21], [44], [24], [25] manage to seek a new representation Z to capture the intrinsic multi-class structure underlying the data:

$$\begin{aligned} &\min_{Z,E} \ \operatorname{rank}(Z) + \lambda\|E\|_1, \\ &\text{s.t. } X = AZ + E, \end{aligned} \tag{1}$$

in which $\operatorname{rank}(\cdot)$ denotes the rank operator for a matrix. $A \in \mathbb{R}^{d\times m}$ is generally defined as the low-rank dictionary with m atoms. $Z \in \mathbb{R}^{m\times n}$ means the newly learned low-rank coefficients and $E \in \mathbb{R}^{d\times n}$ denotes the error component via l_1-norm constraint, targeting at handling noisy data. λ is the balance parameter between two parts.

In general, low-rank representation Z uncovers class structure underlying the cross-view data X by detecting the noise with sparse term. Conventional low-rank models can work well when there is only one dominant factor within the data structure, i.e., class structure [21], [25]. However, it is hard for Z to discover the class structure of cross-view data, since the cross-view divergence within one class is very large. It is easy to notice that there exist more than one factors dominating the data structures, and thus it is challenging to guarantee Z to be low-rank any more and the recovered Z cannot uncover the correlation of samples within one class. To this end, we develop the dual low-rank decompositions to capture the structures of two factors within multi-view data. These two structures are not spanned by each other, since these two structures are intertwined together, which would make the rank of Z larger than c.

As we discussed before, for cross-view data, both identity and view variations would dominate the data distribution. Hence, there exist two independent manifold structures mixed with each other, that is to say, each data sample belongs to two intertwined manifolds. Specifically, both manifolds should be low-rank in terms of two different tasks, since class manifold attempts to capture global class structure, while view manifold aims to uncover the view structure across different classes (Figure 2). To this end, Z in (Eq. (1)) can be divided to two low-rank matrices:

$$\begin{aligned} &\min_{Z_c,Z_v,E} \operatorname{rank}(Z_c) + \operatorname{rank}(Z_v) + \lambda\|E\|_1, \\ &\text{s.t. } X = CZ_c + VZ_v + E, \end{aligned} \tag{2}$$

in which $Z_c \in \mathbb{R}^{m_c\times n}$ and $Z_v \in \mathbb{R}^{m_v\times n}$ denote the low-rank representations for class and view manifolds, respectively. $C \in \mathbb{R}^{d\times m_c}$ and $V \in \mathbb{R}^{d\times m_v}$ are the dictionaries for class structure and view structure with m_c and m_v atoms, respectively. In the ideal case, the rank of Z_c should be c while the rank of Z_v should be k. Actually for cross-view classification, we manage to recover the c-class structure. However, the conventional low-rank representation (Eq. (1)) cannot uncover the c-class structure.

With the objective function (Eq. (2)), two intertwined manifolds can be separated from one another. In this way, Z_c can better represent the global class structure by removing the view structure Z_v. So far, however, an unsupervised decomposition strategy is exploited to separate the two manifold structures, which would be not in the way we are expecting.

B. Discriminative Cross-View Alignment

To effectively supervise the dual low-rank decompositions for the previous approach (Eq. (2)), we propose two discriminative manifold terms to strip down those two intertwined manifolds in a supervised fashion. On the other hand, we are targeting at learning a robust low-dimensional subspace $P \in \mathbb{R}^{d\times p}$ ($p \ll d$) to handle the dimensionality curse. Specifically, along with the recent low-rank subspace learning methods [25], [44], [24], [13], we present our robust view-invariant subspace learning model in the following:

$$\begin{aligned} &\min_{P,Z_c,Z_v,E} \operatorname{rank}(Z_c) + \operatorname{rank}(Z_v) + \lambda\|E\|_1 + \alpha\mathcal{G}(P, Z_c, Z_v) \\ &\text{s.t. } X = CZ_c + VZ_v + E, \ P^\top P = \mathrm{I}_\mathrm{p}, \end{aligned} \tag{3}$$

in which α is the trade-off for the newly designed manifold regularizer $\mathcal{G}(P, Z_c, Z_v)$. Note that the orthogonal constraint $P^\top P = \mathrm{I}_\mathrm{p}$ ($\mathrm{I}_\mathrm{p} \in \mathbb{R}^{p\times p}$) is imposed to avoid invalid solutions.

To make the graph regularizer more effective, we involve both class information and view information as supervised knowledge. Specifically, we explore graph embedding technique and define two graphs, one for class manifolds and the other for view manifolds (Figure 3). To preserve more discriminative knowledge, we attempt to enforce within-class features $K_c = P^\top CZ_c$ ($K_c \in \mathbb{R}^{p\times n}$) more compact while keeping the new low-dimensional within-view features $K_v = P^\top VZ_v$ ($K_v \in \mathbb{R}^{p\times n}$) far away. Thus, we design two novel graph regularizers as follows:

$$\begin{cases} \mathcal{G}_c = \sum\limits_{i,j} \|K_{c,i} - K_{c,j}\|_2^2 W^c_{i,j}, \\ \mathcal{G}_v = \sum\limits_{i,j} \|K_{v,i} - K_{v,j}\|_2^2 W^v_{i,j}, \end{cases}$$

in which $K_{c,i}, K_{c,j}$ are the i-th and j-th column of K_c, while $K_{v,i}, K_{v,j}$ are the i-th and j-th column of K_v. W^c and W^v

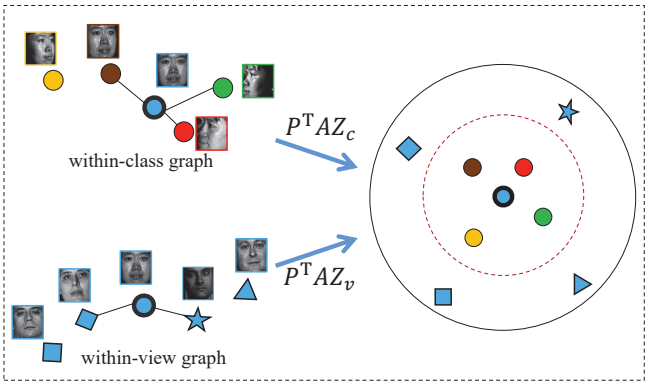


Fig. 3. Illustration of our discriminative cross-view alignment, in which two graphs are built to capture within-class and within-view structures so that it can better supervise the dual low-rank decompositions. Finally, the view variance could be mitigated and a more discriminative low-dimensional subspace would be achieved (shown in right).

represent the weight matrices of two graphs with each element defined in the following way:

$$W^c_{i,j} = \begin{cases} 1, & \text{if } x_i \in \mathcal{C}_{k_1}(x_j), \text{ and } y_i = y_j, \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

$$W^v_{i,j} = \begin{cases} 1, & \text{if } x_i \in \mathcal{V}_{k_2}(x_j), \text{ but } y_i \neq y_j, \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $x_i \in \mathcal{C}_{k_1}(x_j)$ means x_i is the k_1 nearest neighbor of the same class data x_j, while $x_i \in \mathcal{V}_{k_2}(x_j)$ denotes x_i is in the k_2 nearest neighbor of data point x_j underlying the same view. Hence, our two graphs are able to capture the local manifold structure of the within-class data while mitigating the impact of view manifold structure.

Finally, to reduce the intra-class variance while maximizing the margin of inter-class data samples lying in the same view, we develop the graph regularizer $\mathcal{G}(P) = \mathcal{G}(P, Z_c, Z_v)$ and formulate it as:

$$\mathcal{G}(P, Z_c, Z_v) = \frac{\mathcal{G}_c}{\mathcal{G}_v} = \frac{\mathrm{tr}(P^\top C Z_c L_c (P^\top C Z_c)^\top)}{\mathrm{tr}(P^\top V Z_v L_v (P^\top V Z_v)^\top)}, \tag{6}$$

in which $L_{c/v}$ denotes the graph Laplacian for $W^{c/v}$. To make the optimization of Eq. (3) easier, we transform the trace ratio problem to trace difference problem[45], [46] and obtain that:

$$\mathcal{G}(P, Z_c, Z_v) = \mathrm{tr}(P^\top (C Z_c L_c Z_c^\top C^\top - \beta V Z_v L_v Z_v^\top V^\top) P),$$

where β is the trace ratio value of $\mathcal{G}_c$ and $\mathcal{G}_v$, which is automatically updated with $\beta = \frac{\mathrm{tr}(P^\top C Z_c L_c (P^\top C Z_c)^\top)}{\mathrm{tr}(P^\top V Z_v L_v (P^\top V Z_v)^\top)}$ [45], [46].

C. Exponential Rank Approximation

Nuclear norm $\|\cdot\|_*$ is a widely-used surrogate to address the rank minimization [21], [25], [44], [24], which has been proven as the tightest convex approximation to the rank operation. However, it may not be a valid approximation to the rank operation in practical problems, because the rank operation regards all nonzero singular values to have equal contributions while the nuclear norm treats the nonzero singular values differently, i.e., the larger the singular value is, the more contribution it makes to the approximation. There are many research efforts on designing new rank approximation terms [9], [47].

To approximate the rank operation more closely, we adopt the popular exponential rank approximation [47][1] as:

$$F(Z) = f(\sigma(Z)) = \sum_{i=1}^{m} (1 - \exp(\frac{-\sigma_i(Z)}{\delta})), \tag{7}$$

where $\sigma_i(Z)$ is the i-th singular value of Z and δ measures how much close of the rank approximation to the real rank. In general, we can always safely choose a small value. Hence, we set $\delta = 0.1$ as default for simplicity.

To this end, we build the final objective function by integrating exponential rank approximation into Eq. (3) as follows:

$$\begin{aligned} \min_{P, Z_c, Z_v, E} \; & F(Z_c) + F(Z_v) + \lambda \|E\|_1 + \alpha \mathcal{G}(P, Z_c, Z_v) \\ \text{s.t.} \; & X = C Z_c + V Z_v + E, \; P^\top P = \mathrm{I}_{\mathrm{p}}, \end{aligned} \tag{8}$$

where we adopt original data X to replace the class dictionary C and view dictionary V for simplicity.

D. Optimization

To address the optimization problem (8), we explore the first order Taylor expansion instead of Augmented Lagrange Methods (ALM) [21], [24] by reducing extra variables to avoid some matrix multiplications and matrix inverse. Clearly, we first convert the problem (3) to the Augmented Lagrangian format as:

$$\begin{aligned} \mathcal{L} = & F(Z_c) + F(Z_v) + \lambda \|E\|_1 + \alpha \mathcal{G}(P, Z_c, Z_v) \\ & + \langle Q, X - X(Z_c + Z_v) - E \rangle \\ & + \tfrac{\mu}{2} \|X - X(Z_c + Z_v) - E\|_{\mathrm{F}}^2, \end{aligned} \tag{9}$$

in which Q is the Lagrange multiplier and $\mu > 0$ is the penalty parameter. $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm for a matrix, and $\langle , \rangle$ means the inner product operator for two matrices.

Next, we rewrite Eq. (9) by integrating the last three terms into a quadratic form as $\mathcal{L} = F(Z_c) + F(Z_v) + \lambda \|E\|_1 + h(P, Z_c, Z_v, E, Q, \mu) - \frac{1}{\mu}\|Q\|_{\mathrm{F}}^2$, where $h(P, Z_c, Z_v, E, Q, \mu) = \alpha \mathrm{tr}(P^\top X (Z_c L_c Z_c^\top - \beta Z_v L_v Z_v^\top) X^\top P) + \frac{\mu}{2} \|X - X(Z_c + Z_v) - E + \frac{Q}{\mu}\|_{\mathrm{F}}^2$. Similar to the conventional ALM, variables Z_c, Z_v, P and E in Eq. (9) are hard to be jointly solved, however, they are still solvable individually by treating others as constant when updating one. Finally, we solve each sub-problem individually through approximating h to first order Taylor expansion. We further denote $Z_{c,t}, Z_{v,t}, E_t, P_t$ and Q_t as the optimized solution at time t. Specifically, we can obtain the solution to every sub-problem at time $t+1$ as follows:

[1]Motivations for us to exploit the term are 1) it significantly attenuates the contributions from big singular values, by avoiding the unfair penalization of various singular values; 2) by defining $\frac{\partial f(0)}{\partial z_i} = \lim_{z_i \to 0^+} \frac{1}{\delta} \exp(-\frac{z_i}{\delta})$, f is differentiable and concave in $[0, \infty)$.

Updating Z_c:

$$\begin{aligned} Z_{c,t+1} &= \arg\min_{Z_c} F(Z_c) + h(Z_c, Z_{v,t}, E_t, P_t, Q_t, \mu) \\ &= \arg\min_{Z_c} F(Z_c) + \tfrac{\eta_t\mu_t}{2}\|Z_c - Z_{c,t}\|_{\mathrm{F}}^2 \\ &\qquad + \langle \nabla_{Z_c} h, Z_c - Z_{c,t}\rangle \\ &= \arg\min_{Z_c} \tfrac{1}{\eta_t\mu_t} F(Z_c) + \tfrac{1}{2}\|Z_c - Z_{c,t} + \nabla_{Z_c} h\|_{\mathrm{F}}^2, \end{aligned} \tag{10}$$

where $\nabla_{Z_c} h = \nabla_{Z_c} h(Z_{c,t}, Z_{v,t}, E_t, P_t, Q_t, \mu_t) = 2\alpha X^\top P_t P_t^\top X Z_{c,t} L_c - X^\top Q_t - \mu_t X^\top (X - X(Z_{c,t} + Z_{v,t}) - E_t)$ and $\eta_t = \|X\|_{\mathrm{F}}^2$. This can be addressed by the theorem in Appendix. Specifically, we adopt $\bar{Z} = Z_{c,t} - \nabla_{Z_c} h$ in Eq. (15).

Updating Z_v:

$$\begin{aligned} Z_{v,t+1} &= \arg\min_{Z_v} F(Z_v) + h(Z_{c,t+1}, Z_v, E_t, P_t, Q_t, \mu_t) \\ &= \arg\min_{Z_v} F(Z_v) + \tfrac{\eta_t\mu_t}{2}\|Z_v - Z_{v,t}\|_{\mathrm{F}}^2 \\ &\qquad + \langle \nabla_{Z_v} h, Z_v - Z_{v,t}\rangle \\ &= \arg\min_{Z_v} \tfrac{1}{\eta_t\mu_t} F(Z_v) + \tfrac{1}{2}\|Z_v - Z_{v,t} + \nabla_{Z_v} h\|_{\mathrm{F}}^2, \end{aligned} \tag{11}$$

in which $\nabla_{Z_v} h = \nabla_{Z_v} h(Z_{c,t+1}, Z_{v,t}, E_t, P_t, Q_t, \mu_t) = -2\alpha X^\top P_t P_t^\top X Z_{v,t} L_v - X^\top Q_t - \mu_t X^\top (X - X(Z_{c,t+1} + Z_{v,t}) - E_t)$. Problem (11) can be addressed in the same manner with Eq. (10). Specifically, we use $\bar{Z} = Z_{v,t} - \nabla_{Z_v} h$ in Eq. (15).

Updating E:

$$E_{t+1} = \arg\min_{E} \frac{\lambda}{\mu_t}\|E\|_1 + \frac{1}{2}\|E - (\widetilde{X}_{t+1} + \frac{Q_t}{\mu_t})\|_{\mathrm{F}}^2, \tag{12}$$

in which we denote $\widetilde{X}_{t+1} = X - X(Z_{c,t+1} + Z_{v,t+1})$ for simplicity. Eq. (12) can be solved by defining $\bar{E} = \widetilde{X}_{t+1} + \dfrac{Q_t}{\mu_t}$. Specifically, we have $E = E_\mu(\bar{E})$, which is denoted component-wisely as $[E_\mu(\bar{E})]_{ij} = \max\{|\bar{e}_{ij}| - \dfrac{\lambda}{\mu_t}, 0\}\mathrm{sign}(\bar{e}_{ij})$.

Updating P:

$$P_{t+1} = \arg\min_{P^\top P = \mathrm{I}_{\mathrm{p}}} \alpha \mathrm{tr}(P^\top X \widetilde{Z}_{t+1} X^\top P), \tag{13}$$

where we define $\widetilde{Z}_{t+1} = Z_{c,t+1} L_c Z_{c,t+1}^\top - \beta Z_{v,t+1} L_v Z_{v,t+1}^\top$ for simplicity. Eq. (13) is a standard graph embedding objective function, which can be solved by

$$\alpha X \widetilde{Z}_{t+1} X^\top \rho = \xi \rho. \tag{14}$$

It is easy to demonstrate that $\alpha X \widetilde{Z}_{t+1} X^\top$ is symmetric and positive semidefinite by given β. The vectors $\rho_i (i = 0, 1, \cdots, p-1)$ that minimize the objective function are according to the minimum eigenvalue solutions for the Eigen-decomposition problem. Thus, we could achieve the linear projection as $P = [\rho_0, \cdots, \rho_{p-1}]$.

To sum up, we list the detailed solutions to Eq. (9) in **Algorithm 1**, in which we empirically set μ_0, ρ, ϵ, $t_{\max}$ and $\mu_{\max}$, while tuning two other parameters (λ and α) during the experiment. We initialize P with a random matrix, $Z_{c/v}$ with Eq. (1).

Algorithm 1 Optimization to Eq. (8)

Input: X, $\lambda, \alpha, L_c, L_v$

Initialize: $E_0 = Q_0 = 0, \epsilon = 10^{-6}, \rho = 1.3, \mu = 10^{-6}, \mu_{\max} = 10^6, t_{\max} = 10^3, \beta = 1, t = 0.$

while not converged **or** $t \leq t_{\max}$ **do**

1. Update $Z_{c,t+1}$ through Eq. (10) by fixing others;
2. Update $Z_{v,t+1}$ through Eq. (11) by fixing others;
3. Update E_{t+1} through Eq. (12) by fixing others;
4. Update P_{t+1} through Eq. (14) by fixing others,
5. Update the multiplier Q_{t+1}:
   $Q_{t+1} = Q_t + \mu(\widetilde{X}_{t+1} + E_{t+1})$;
6. Update the parameter μ and β by
   $\beta = \dfrac{\mathrm{tr}(P^\top X Z_c L_c (P^\top X Z_c)^\top)}{\mathrm{tr}(P^\top X Z_v L_v (P^\top X Z_v)^\top)}$;
   $\mu = \min(\rho\mu, \mu_{\max})$;
7. Check the convergence conditions
   $\|\widetilde{X}_{t+1} + E_{t+1}\|_\infty < \epsilon$.
8. $t = t + 1$.

end while

output: Z_c, Z_v, E, P

E. Computational Analysis

To make it simple, we mainly analyze the optimization complexity presented in **Algorithm 1**. $X \in \mathbb{R}^{d\times n}$, $P \in \mathbb{R}^{d\times p}$ and $Z_c \in \mathbb{R}^{n\times n}, Z_v \in \mathbb{R}^{n\times n}$. Through our optimization, we find the most consuming parts include the rank optimization in Step 1& 2, and Eigen-decomposition in Step 4.

Traditional SVD operation in Step 1&2 would take $\mathcal{O}(n^3)$ for Z_c, Z_v, repetitively. When the sample size n becomes larger, fortunately, we could further accelerate Step 1& 2 to $\mathcal{O}(r_{c/v}^2 n)$, in which $r_{c/v}$ is the rank of C/V by the recent fast low-rank method [21]. So RCVL is quite scalable for large-scale datasets, given low-rank dictionaries C/V. When we use X to replace C/V, the computational cost would $\mathcal{O}(d^2 n)$ at most (suppose $d \leq n$). This is also efficient given that the data dimension d is not too high. Moreover, we adopt divide-conquer strategy to further fasten our solution optimization [48], [49]. Specifically, we randomly sample data from the training set. For each subset $\bar{X}_i \in \mathbb{R}^{d\times n_i}$ with its two sub-graphs, we have the optimized $\bar{Z}_{c,i}, \bar{Z}_{v,i}$, then fuse multiple small low-rank matrices to achieve Z_c, Z_v [48], [49]. On the other hand, eigen-decomposition in Step 4 on matrix with size $d \times d$ costs close to $\mathcal{O}(d^3)$, which could be further reduced to $O(d^{2.376})$ through the Coppersmith-Winograd theorem [50].

All in all, the complexity for the proposed approach is $\mathcal{O}\Big(t\big(d^{2.376} + 2d^2 n\big)\Big)$.

IV. Experimental Results

In this part, we use several cross-domain benchmarks to evaluate our proposed algorithm. First of all, we show the datasets' details and experimental protocols. Following that, we present the comparison results, properties analysis and discussion.

A. Datasets & Experimental Setting

Five cross-view image datasets, i.e., CMU-PIE face, Extended Yale B face, MS-Celeb-1M face datasets, COIL-100 object and ALOI-100 object datasets are evaluated in our experiment.

CMU-PIE Face database[2] includes 68 individuals in all. Examples for each individual are under 21 various illumination conditions. We conduct several rounds of experiments by changing the size of poses from two to five to build different evaluations. Face images are cropped into 64×64 and the raw features are used. To further evaluate the robustness of all comparisons, we artificially corrupt the images with 10% random noise. We also add block noise with random corruption. Specifically, we randomly add a 20×20 block to the original images.

Extended Yale Face Database B[3] consists of 16,128 images from 28 individuals under 64 lighting conditions and 9 viewpoints. We also conduct several times of experiments by changing the size of poses from two to five to construct different evaluations. Per pose, we randomly select 10 images to construct the training set, whilst the remaining images are used for test. We crop images into 270×250 and adopt the raw features with PCA preprocessing to 3,000 dimensions. Since 9 poses are all near frontal poses, we add 20%, 40% and 60% random noise for each image to further evaluate the robustness of all comparisons.

MS-Celeb-1M[4] is a real-world large-scale face database covering the top 100K celebrities. For each celebrity, public search engines are used to obtain approximately 100 images, leading to around 10M images. Same celebrity may show very large divergence in different images. In this experiment, we further select the celebrity with more than 50 images. In total, we have 8,172 individuals with 644,748 images. We adopt two most recent deep structures to extract the feature representations, i.e., VGG-face [52] with dimension as 4,096 and Center-face [17] with dimensions as 1,024.

COIL-100 object database[5] contains 100 categories. Images per object were captured 5 degrees apart and thus, each object contains 72 samples. Following [24], we partition COIL-100 into two parts: "COIL1" and "COIL2". COIL1 includes images from $[0^o, 85^o] \cup [180^o, 265^o]$ and COIL2 consists of images from $[90^o, 175^o] \cup [270^o, 355^o]$. The pixel values with 20% corruption out of 64×64 are input as the features.

ALOI object database[6] includes 1,000 categories, where we select the first 100 categories with 7,200 images to build ALOI-100. The same to COIL-100, ALOI-100 contains 72 samples per object. Thus, we apply the same setting with COIL-100 to 4 views of ALOI-100. The pixel values with 20% corruption out of 96×72 are input as the features.

[2]http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html
[3]http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html
[4]https://www.msceleb.org/
[5]http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php
[6]http://aloi.science.uva.nl/

B. Comparison Results

In our experiments, we attempt to solve the challenging cross-view problem, where we only know the view knowledge of training data, while we do not access to the view knowledge of the test data [24], [13]. Aiming to testify the effectiveness of our proposed models, we further compare with traditional multi-view learning approaches, i.e, CMML [1], JFSSL [6], MvDA [5] and MvDA-VC [5], where we provide the extra view information of the test data. Actually, CMML is designed for two-view cases, thus, we only show the results of two views. While JFSSL, MvDA and MvDA-VC are designed for multiple views by seeking a view-invariant space. To further demonstrate the effectiveness of our model, we compare with two most recent deep CNN face extractors, i.e., VGG-face [52] and Center-face [17]. The goal is to demonstrate that deep CNN cannot perfectly handle the view variance although it is trained on a large-scale dataset.

Moreover, we compare with recent robust feature extraction algorithms, i.e., LatLRR [51], SRRS [25], LRCS [24] and our conference version (RMSL) [26]. Specifically, SRRS, RMSL, and our proposed approach are supervised; LatLRR is totally unsupervised; while LRCS is a weakly supervised method, since it only needs to access the view information of the data. RMSL is our previous conference version, which is denoted as **Ours-I**, while our current version is named as **Ours-II**. For both our models, we simply set $k_1 = 5$ and $k_2 = 10$ across all the datasets. For subspace learning based models, it is also an important parameter p, which we simply set as 100 for all the datasets. For all the comparisons, we provide the classification accuracy through the nearest neighbor classifier (1-NNC).

To CMU-PIE face database, we choose 10 samples per individual each pose randomly to construct the training set, while the remaining face samples are adopted to evaluate all the algorithms. In all, we randomly select five evaluations and obtain the average accuracy. Tables I, II & III show the recognition performance of 9 algorithms on original, randomly corrupted and block corrupted face images, in which Case 1: {C02, C14}, Case 2: {C02, C27}, Case 3: {C14, C27}, Case 4: {C05, C29}, Case 5: {C05, C07, C29}, Case 6: {C05, C14, C29, C34}, Case 7: {C02, C05, C14, C29, C31}. Furthermore, we adopt VGG-face [52] and Center-face [17] to extract the deep features from CMU-PIE to evaluate our model based the deep features (shown in Table VIII). With more views involved, JFSSL, MvDA and MvDA-VC can work better than our model in the clean cases.

To Extended Yale B face database, we build more corrupted cases since the pose variance of Yale B face is much smaller than that of CMU-PIE face. Specifically, we randomly add 20%, 40% noise to evaluate different comparisons. For Case 1-3, we evaluate two-view case {Y07,Y08} with 20%, 40% and 60%, respectively. For Case 4, three-view combination {Y02,Y03,Y05} with 20% corruption is used for evaluation. For Case 5&6, we adopt {Y01,Y04,Y06} three views with 20% and 40% corruption, respectively. For Case 7, we use {Y01,Y04,Y06,Y09} four views with 20% corruption. Table IV shows the comparison performance of all algorithms on 7 cases.

TABLE I
COMPARISON RESULTS (%) OF 9 ALGORITHMS ON THE **ORIGINAL** CMU-PIE MULTI-POSE FACE DATABASE.

Algorithms	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
LatLRR [51]	77.92±0.13	76.24±0.12	75.29±0.17	83.68±0.17	69.74±0.15	42.54±0.12	35.33±0.14
SRRS [25]	78.27±0.04	78.74±0.23	77.45±0.02	86.28±0.09	71.44±0.13	43.86±0.12	35.66±0.12
LRCS [24]	87.78±0.12	86.67±0.11	87.38±0.19	89.12±0.12	74.84±0.14	44.48±0.13	36.17±0.11
CMML [1]	86.23±0.11	86.98±0.16	87.79±0.20	90.65±0.14	-	-	-
JFSSL [6]	87.83±0.12	87.86±0.10	88.15±0.16	92.48±0.12	73.87±0.16	52.18±0.18	46.66±0.12
MvDA [5]	86.76±0.15	86.12±0.12	86.92±0.12	91.23±0.10	72.46±0.14	50.04±0.15	45.36±0.14
MvDA-VC [5]	87.82±0.12	87.81±0.09	88.18±0.13	92.43±0.12	75.36±0.18	**54.13±0.16**	**47.67±0.18**
RMSL (Ours-I) [26]	89.15±0.06	88.05±0.07	88.40±0.17	93.95±0.11	75.16±0.12	44.93±0.11	37.14±0.08
Ours-II	**89.47±0.12**	**88.26±0.13**	**89.24±0.19**	**94.98±0.15**	**75.84±0.14**	45.88±0.15	37.93±0.10

TABLE II
COMPARISON RESULTS (%) OF 9 ALGORITHMS ON **CORRUPTED** CMU-PIE MULTI-POSE FACE DATABASE.

Algorithms	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
LatLRR [51]	73.10±0.07	73.24±0.32	73.85±0.12	75.21±0.08	58.94±0.09	39.26±0.12	32.07±0.23
SRRS [25]	72.27±0.15	72.74±0.18	71.45±0.08	74.19±0.13	54.32±0.13	39.34±0.22	32.03±0.22
LRCS [24]	78.98±0.13	78.67±0.15	78.38±0.26	80.54±0.12	65.84±0.24	39.48±0.23	32.57±0.21
CMML [1]	74.83±0.09	73.48±0.21	75.94±0.15	75.76±0.16	-	-	-
JFSSL [6]	77.98±0.12	75.25±0.19	77.92±0.13	78.72±0.19	67.98±0.12	**45.19±0.14**	35.17±0.19
MvDA [5]	75.34±0.11	74.81±0.13	76.36±0.16	77.28±0.18	65.43±0.14	44.20±0.15	34.68±0.21
MvDA-VC [5]	77.26±0.12	75.54±0.09	77.84±0.15	78.98±0.13	66.76±0.17	44.86±0.13	34.96±0.20
RMSL (Ours-I) [26]	82.12±0.18	82.67±0.14	82.38±0.17	84.18±0.12	69.84±0.19	43.87±0.19	35.78±0.12
Ours-II	**83.45±0.13**	**83.48±0.14**	**83.25±0.16**	**85.23±0.15**	**70.28±0.18**	44.69±0.12	**36.92±0.13**

TABLE III
COMPARISON RESULTS (%) OF 9 ALGORITHMS ON CMU-PIE MULTI-POSE FACE DATABASE WITH BLOCK CORRUPTION.

Algorithms	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
LatLRR [51]	46.51±0.14	46.83±0.16	46.29±0.21	47.91±0.16	46.06±0.18	32.27±0.17	30.21±0.13
SRRS [25]	54.63±0.13	55.38±0.15	54.95±0.18	56.45±0.14	53.26±0.20	35.62±0.13	32.81±0.12
LRCS [24]	58.71±0.16	59.25±0.15	58.87±0.16	59.24±0.13	57.34±0.18	36.83±0.14	33.23±0.16
CMML [1]	56.25±0.14	57.35±0.17	56.36±0.19	56.82±0.15	-	-	-
JFSSL [6]	57.52±0.12	58.48±0.14	57.49±0.16	57.84±0.18	55.82±0.15	39.21±0.14	36.72±0.18
MvDA [5]	58.37±0.16	59.24±0.19	58.17±0.17	58.78±0.17	56.26±0.14	38.76±0.18	36.26±0.14
MvDA-VC [5]	60.43±0.13	60.32±0.13	61.84±0.18	60.91±0.18	57.38±0.16	39.98±0.15	37.85±0.18
RMSL (Ours-I) [26]	64.21±0.18	64.92±0.19	65.86±0.20	65.84±0.19	**63.56±0.17**	**42.78±0.18**	**39.68±0.17**
Ours-II	**64.23±0.17**	**65.25±0.20**	**65.93±0.16**	**66.87±0.18**	63.48±0.15	42.45±0.16	39.20±0.16

To MS-Celeb-1M face dataset, we also exploit VGG-face [52] and Center-face [17] to extract the deep features. Furthermore, we randomly select $s = \{20, 30, 40, 50\}$ percentage as the training data while the remaining as the test data. For this setting, we aim to verify our model can well extend to large-scale dataset and further improve the performance based on deep features. Thus, we only compare our model with the original deep features and the results are shown in Table VII.

To object datasets, we select one subset from COIL1 (ALOI1) and one subset from COIL2 (ALOI2) for training, while the remaining two views as the test set. In total, we can build 4 cases for evaluation. The results are presented in Table V & VI.

Discussion: From the comparison performance, we notice that our proposed algorithm achieves better recognition performance than other comparisons in most cases, except JFSSL, MvDA and MvDA-VC, which are accessible to the view knowledge of the test data. This verifies that our designed approach is an effective compromise when we have no view information for the test data in reality. As we can see, JFSSL/MvDA/MvDA-VC have a superiority when more views are involved, as multiple view-specific projections have a stronger ability to align each specific view. However, JFSSL/MvDA/MvDA-VC belong to conventional multi-view subspace learning, which is sensitive to the corruption. Differently, our proposed approach integrates low-rank modeling and subspace learning into a unified framework, which tends to well fight off the corruption in real applications. When the corruption may not bring in more influence than the original view variances (case 6 in Table II), learning multiple view-specific projections (e.g., JFSSL, MvDA, MvDVAC) would better handle this challenge than one common projection (e.g., Ours).

Specifically, to CMU-PIE face database, we notice that all the approaches cannot work well with more views involved, since the within-class variance becomes much larger. While for the four 2-view evaluations, all the approaches obtain very similar performance, as we consider that the view divergence for all 2-view cases is almost the same to some extent. For these four 2-view evaluations, our model shows

TABLE IV
COMPARISON RESULTS (%) OF 9 ALGORITHMS ON THE EXTENDED YALE B FACE DATABASE.

Algorithms	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
LatLRR [51]	88.34±0.17	83.24±0.19	75.31±0.18	89.46±0.21	78.74±0.23	76.82±0.20	79.93±0.20
SRRS [25]	92.46±0.19	89.29±0.20	83.21±0.18	92.87±0.20	84.56±0.24	82.78±0.19	85.02±0.21
LRCS [24]	90.54±0.20	84.34±0.19	76.62±0.23	91.21±0.22	80.58±0.21	79.49±0.20	81.25±0.17
CMML [1]	93.18±0.21	90.47±0.20	82.28±0.20	-	-	-	
JFSSL [6]	94.26±0.22	91.82±0.26	83.76±0.18	93.06±0.20	87.18±0.20	86.32±0.23	87.26±0.19
MvDA [5]	94.08±0.17	91.62±0.23	83.85±0.19	93.26±0.18	87.32±0.18	86.48±0.22	87.62±0.18
MvDA-VC [5]	94.42±0.18	92.38±0.21	84.72±0.17	93.73±0.21	87.96±0.22	86.86±0.23	87.79±0.15
RMSL (Ours-I) [26]	95.16±0.23	**93.54**±0.26	83.94±0.16	**94.23±0.22**	87.54±0.27	85.58±0.24	87.98±0.20
Ours-II	**95.81**±0.21	92.01±0.23	**85.52**±0.19	93.42±0.23	**88.90±0.20**	**87.16±0.21**	**88.67**±0.17

TABLE V
RECOGNITION RESULTS OF 9 ALGORITHMS ON 4 CASES OF THE 20% CORRUPTED COIL-100 DATASET, WHERE CASE 1: VIEW 1 AND VIEW 3; CASE 2: VIEW 1 AND VIEW 4; CASE 3: VIEW 2 AND VIEW 3; CASE 4: VIEW 2 AND VIEW 4.

Methods	LatLRR[51]	SRRS[25]	LRCS [24]	CMML [1]	JFSSL [6]	MvDA [5]	MvDA-VC [5]	RMSL (Ours-I) [26]	Ours-II
Case 1	71.09	75.44	72.89	76.98	77.22	77.64	77.54	77.97	**78.58**
Case 2	73.24	78.23	76.06	80.24	80.65	80.87	80.96	80.32	**82.16**
Case 3	75.43	79.25	77.19	80.76	81.24	80.96	81.14	81.32	**82.97**
Case 4	70.98	72.97	71.75	76.43	77.24	77.65	77.98	77.41	**79.13**

TABLE VI
RECOGNITION RESULTS OF 9 ALGORITHMS ON 4 CASES OF THE 20% CORRUPTED ALOI-100 DATASET, WHERE CASE 1: VIEW 1 AND VIEW 3; CASE 2: VIEW 1 AND VIEW 4; CASE 3: VIEW 2 AND VIEW 3; CASE 4: VIEW 2 AND VIEW 4.

Methods	LatLRR[51]	SRRS[25]	LRCS [24]	CMML [1]	JFSSL [6]	MvDA [5]	MvDA-VC [5]	RMSL (Ours-I) [26]	Ours-II
Case 1	76.32	80.44	79.36	79.28	80.14	80.08	80.28	81.23	**82.98**
Case 2	72.12	76.92	74.78	76.23	76.76	76.84	77.36	77.45	**78.23**
Case 3	69.32	74.42	74.33	75.26	75.86	75.84	75.94	75.32	**76.82**
Case 4	75.64	79.11	77.53	79.62	80.24	80.16	80.48	80.23	**81.85**

perfect superiority over other comparisons, which represents our view-invariant subspace well uncovers the intrinsic structure underlying 2-view face images. However, for 3-view combination, our proposed approach is not able to achieve a large improvement, since these three views show a lower view divergence.

Since the pose variance of Extended Yale B face database is not very large comparing with CMU-PIE. We could observe that the performance of all comparisons may increase when more views are involved as more views could help capture the intrinsic structures of each class. However, we find with more artificial noise involved into the data, our algorithm is very robust while traditional multi-view learning algorithms degrade a lot (Seen from Table 3 and Tables 4&5). Similar to Extended Yale B face, two object databases also show small view variance so that our algorithm could only obtain a little bit better performance.

MS-Celeb-1M involves many view variances in the wild, which is much more challenging although adopting the deep features. However, our model would further improves the performance based on the effective deep features with 3-4% margins (Table VII). This verifies the effectiveness of our model in large-scale real-world dataset. Furthermore, our proposed model could further improve the performance based on the deep features, as the results shown in Table VIII. This phenomenon verifies the effectiveness of our proposed model.

Finally, we further conduct t-test to demonstrate the statistical significance of our approach, whose p-value results of our method by comparing with others are provided in Figure 4. The performance comparison of two approaches is statistically significant when the p-value is smaller than 0.05. To make easy observation, we perform $-\log(p)$ pre-processing, that is to say, the difference of two algorithms is significantly when the values are larger than $-\log(0.05) \approx 2.9$.

TABLE VII
COMPARISON RESULTS (%) ON MS-CELEB-1M FACE DATABASE.

Algorithms	s = 20	s = 30	s = 40	s = 50
VGG-face [52]	62.31	75.70	85.44	92.25
VGG-face [52]+Ours-II	66.24	78.46	87.92	94.12
Center-face [17]	55.99	58.88	74.39	86.58
Center-face [17]+Ours-II	60.25	62.23	77.27	88.28

C. Empirical Evaluation

In this part, we analyze some properties of our method, e.g., robustness, convergence analysis, parameter influence as well as computation cost.

First of all, we show the influences of various corruption ratios to various comparisons and evaluate on 0%, 10%, 20%, 30%, 40%, and 50% corruptions with two-view case from CMU-PIE face dataset, i.e., {C02,C14} (Fig. 5 (a)). We also use 10 images per pose for training while left as test, and present the recognition results in Figure 5 (a), in which our proposed approach in two modes consistently performs better than others. This verifies that our proposed approach is able to learn a more robust feature extractor, especially when data

TABLE VIII
COMPARISON RESULTS (%) ON THE DEEP FEATURES OF THE CMU-PIE FACE DATABASE.

Algorithms	{C14,C22}	{C05,C14}	{C05,C22}	{C14,C34}	{C09,C27}	{C05,C14,C22}	{C05,C14,C22,C34}
VGG-face [52]	94.28	93.88	96.85	92.97	92.41	85.43	80.76
VGG-face [52]+Ours-II	95.17	94.41	97.23	94.35	92.52	86.28	82.01
Center-face [17]	92.34	91.53	92.67	90.33	87.44	78.46	74.79
Center-face [17]+Ours-II	93.68	92.63	94.25	91.43	88.24	79.14	75.27

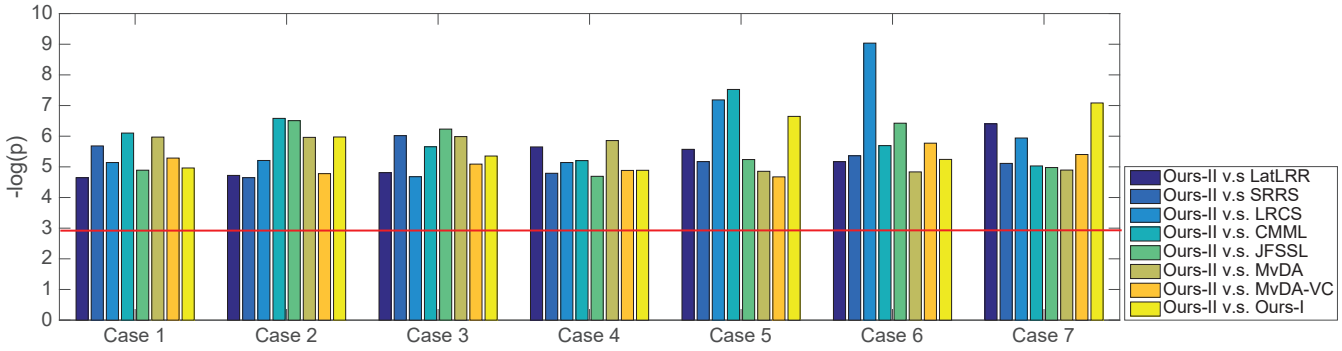
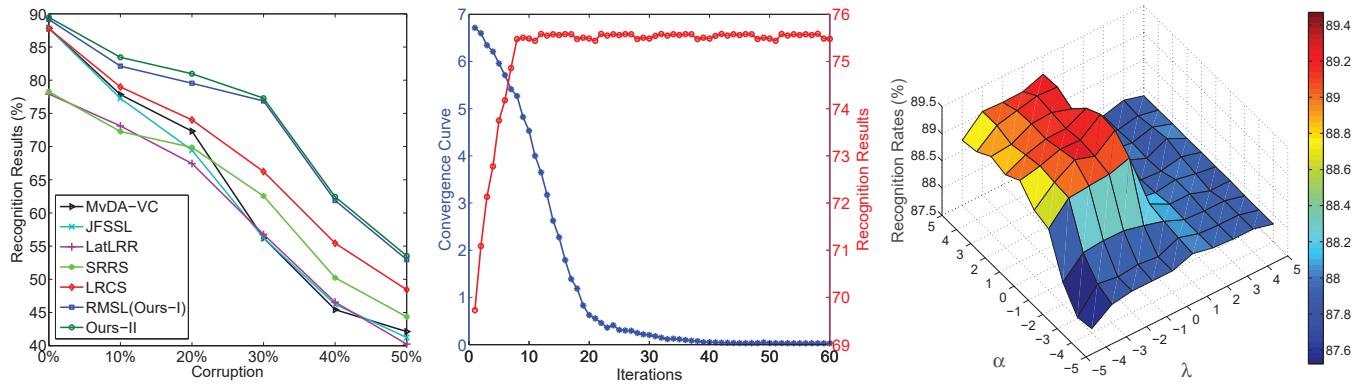


Fig. 4. p-value of t-test between our method and others on the **original** CMU-PIE multi-pose face database. We do pre-processing using $-\log(p)$ so that the large value shown in the figure means the more significance of one method compared with the other.



(a) (b) (c)

Fig. 5. (a) Robustness evaluation of all comparisons on two-view case from CMU-PIE face database. (b) convergence curve (Blue '*') and recognition curve (red 'o') of our algorithm for one selection in Case 5 {C05,C09,C27} of CMU-PIE face database. (c) Recognition rate of our algorithm with different parameter values $\{\alpha, \lambda\}$ on Case 2 {C02,C27} of CMU-PIE face database. The values of x-axis and y-axis are used $\log()$ to rescale the length.

are largely corrupted. That is, our approach is more efficient in real-world scenarios under different noisy conditions.

Secondly, we evaluate the convergence of **Algorithm 1** to empirically show the convergence through experiments in different iterations, which is calculated via the relative error: $\|X - A(Z_c + Z_v) - E\|_{\mathrm{F}}^2 / \|X\|_{\mathrm{F}}^2$. Note that we divide the convergence condition in **Algorithm 1** with the data scale, which would make the curve more smooth and fine. Specifically, we evaluate on 3-view case on CMU-PIE face dataset. The convergence curve of the proposed approach is reported in Figure 5 (b), as well as the classification performance. We observe our approach converges pretty well and efficiently. Besides, we witness that the classification performance goes up very quickly and after that keeps at a stable value. We notice that the performance (recognition results) is stable (constant) even if the convergence is not established. We consider that we could easily achieve very optimal projection P with good performance, since we adopt a supervised graph regularizer to guide the projection. However, we may take more iterations to optimize $Z_{c/v}$ and E.

TABLE IX
TRAINING TIME OF FOUR ALGORITHMS ON CMU-PIE FACE DATASET.

Config	2 Views	3 Views	4 Views	5 Views
LatLRR [51]	291.5	817.7	1635.4	2736.9
LRCS[24]	184.0	547.3	1305.3	2311.1
RMSL (Ours-I)[26]	72.3	162.6	311.3	510.4
Ours-II	74.1	166.8	313.2	513.6

Thirdly, we evaluate parameter influence for our newly designed model (Ours-II). For better illustration, we simultaneously analyze two parameters on two-view case, i.e., {C02,C27}, from CMU-PIE face database. Parameter analysis are reported in Figure 5(c). We notice that larger value of α shows better performance, especially for λ with small values. On the other hand, we observe λ around 10^{-1} provides relatively better results. Thus, we set $\alpha = 10^2$ and $\lambda = 10^{-1}$ throughout the experiments.

Finally, we also testify the training cost for our approach through comparing several other models. Specifically, we evaluate on different cases of CMU-PIE face dataset, and we report the training time for all comparisons with 10 iterations. We conduct experiments on Matlab 2014b, CPU i7-3770 and 32 GB memory size. The computational time for 4 algorithms are reported in Table IX (unit is *second*). From the results, our proposed approach is more efficient than LRCS and LatLRR. This mainly attributes to our proposed efficient solution, avoiding extra variables.

V. Conclusion

In this paper, we proposed a Robust Cross-view Subspace Learning algorithm to build a view-free projection to alleviate cross-view data analysis. In details, we developed a dual low-rank decompositions to unweave two intertwined manifolds, and thus our algorithm could preserve more class-wise knowledge for better classification by mitigating the impact from view divergence under the intra-class data. Moreover, two discriminative graphs were incorporated into our dual low-rank decompositions to make it more effective. Experiments on several face and object benchmarks demonstrated the superiority of our proposed approach, compared with the low-rank based methods and state-of-the art deep face models.

Acknowledgment

This work is supported in part by the NSF IIS award 1651902, NIJ Graduate Research Fellowship 2016-R2-CX-0013, ONR Young Investigator Award N00014-14-1-0484, and U.S. Army Research Office Award W911NF-17-1-0367.

Appendix

Since $F(Z) = f(\sigma(Z))$ is a unitarily invariant function, updating $Z_{c/v}$ can be formulated as the following optimization problem:

$$\min_{Z} F(Z) + \frac{\mu}{2}\|Z - \bar{Z}\|_{\mathrm{F}}^2, \tag{15}$$

where $\mu > 0$ and $\bar{Z} \in \mathbb{R}^{n\times n}$ with its SVD as $U\Sigma_{\bar{Z}}V^\top$ ($\Sigma_{\bar{Z}} = \mathrm{diag}(\sigma_{\bar{Z}})$).

The optimal solution Z^* to Eq. (15) can be achieved by the Moreau-Yosida operator $\sigma^* = \mathrm{prox}_{f,\mu}(\sigma_Z)$. Specifically, Z^* is with SVD being $U\Sigma_Z^*V^\top$, where $\Sigma_Z^* = \mathrm{diag}(\sigma_Z^*)$. Next, we have

$$\mathrm{prox}_{f,\mu}(\sigma_Z) = \arg\min_{\sigma\geq 0} f(\sigma) + \frac{\mu}{2}\|\sigma - \sigma_Z\|_2^2.$$

In this case, we can resort to the difference of convex (DC) [53] strategy, since the first term is concave while the second one is convex to σ. We apply a linear approximation at each iteration of DC programming. At the $(\tau+1)$-th iteration, we achieve that

$$\sigma^{\tau+1} = (\sigma_Z - \frac{\partial f(\sigma^\tau)}{\mu_t\eta_t})_+,$$

where $\partial f(\sigma^\tau)$ is the gradient of $f(\cdot)$ at σ^τ.

References

[1] A. Mignon and F. Jurie, "Cmml: a new metric learning approach for cross modal matching," in *Asian Conference on Computer Vision*, 2012, pp. 1–14.

[2] X. Cai, C. Wang, B. Xiao, X. Chen, and J. Zhou, "Regularized latent least square regression for cross pose face recognition," in *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, 2013, pp. 1247–1253.

[3] K. Wang, R. He, W. Wang, L. Wang, and T. Tan, "Learning coupled feature spaces for cross-modal matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2088–2095.

[4] J. Wang, X. Nie, Y. Xia, Y. Wu, and S.-C. Zhu, "Cross-view action modeling, learning and recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2649–2656.

[5] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 188–194, 2016.

[6] K. Wang, R. He, L. Wang, W. Wang, and T. Tan, "Joint feature selection and subspace learning for cross-modal retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 10, pp. 2010–2023, 2016.

[7] L. Lin, G. Wang, W. Zuo, X. Feng, and L. Zhang, "Cross-domain visual matching via generalized similarity measure and feature learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1089–1102, 2017.

[8] S. Wang, Z. Ding, and Y. Fu, "Coupled marginalized auto-encoders for cross-domain multi-view learning," in *Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press, 2016, pp. 2125–2131.

[9] Z. Ding, M. Shao, and Y. Fu, "Low-rank embedded ensemble semantic dictionary for zero-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2050–2058.

[10] Z. Tao, H. Liu, S. Li, Z. Ding, and Y. Fu, "From ensemble clustering to multi-view clustering," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 2017, pp. 2843–2849.

[11] M. Brbić and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognition*, vol. 73, pp. 247–258, 2018.

[12] Z. Ding, M. Shao, and Y. Fu, "Robust multi-view representation: A unified perspective from multi-view learning to domain adaption." in *27th International Joint Conference on Artificial Intelligence*, 2018, pp. 5434–5440.

[13] Z. Ding and Y. Fu, "Robust multiview data analysis through collective low-rank subspace," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 5, pp. 1986–1997, 2018.

[14] H. Zhao, H. Liu, Z. Ding, and Y. Fu, "Consensus regularized multi-view outlier detection," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 236–248, 2018.

[15] H. Zhao, Z. Ding, and Y. Fu, "Pose-dependent low-rank embedding for head pose estimation," in *the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI Press, 2016, pp. 1422–1428.

[16] Y. Kong, Z. Ding, J. Li, and Y. Fu, "Deeply learned view-invariant features for cross-view action recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 3028–3037, 2017.

[17] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 499–515.

[18] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 625–632.

[19] Y. Quan, Y. Xu, Y. Sun, Y. Huang, and H. Ji, "Sparse coding for classification via discrimination ensemble," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5839–5847.

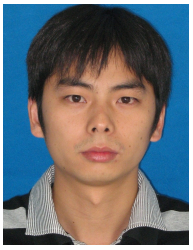
[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, p. 11, 2011.

[21] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 171–184, 2013.

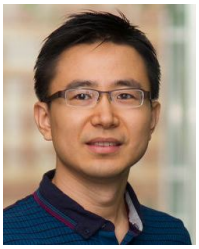
[22] Y. Liu, X. Li, C. Liu, and H. Liu, "Structure-constrained low-rank and partial sparse representation with sample selection for image classification," *Pattern Recognition*, vol. 59, pp. 5–13, 2016.

[23] Y.-X. Wang, H. Xu, and C. Leng, "Provable subspace clustering: When lrr meets ssc," in *Advances in Neural Information Processing Systems*, 2013, pp. 64–72.

[24] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *IEEE International Conference on Data Mining*. IEEE, 2014, pp. 110–119.
[25] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2160–2173, 2016.
[26] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 1181–1187.
[27] B.-K. Bao, G. Liu, R. Hong, S. Yan, and C. Xu, "General subspace learning with corrupted training data via graph embedding," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4380–4393, 2013.
[28] F. Zhang, J. Yang, Y. Tai, and J. Tang, "Double nuclear norm-based matrix decomposition for occluded image recovery and background modeling," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1956–1966, 2015.
[29] X. Jiang and J. Lai, "Sparse and dense hybrid representation via dictionary decomposition for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 5, pp. 1067–1079, 2015.
[30] Y. Panagakis, M. Nicolaou, S. Zafeiriou, and M. Pantic, "Robust correlated and individual component analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1665–1678, 2016.
[31] W. Wang, R. Arora, K. Livescu, and J. Bilmes, "On deep multi-view representation learning," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 1083–1092.
[32] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Conference on Data Mining and Data Warehouses*, 2010, pp. 1–4.
[33] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization." in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 2921–2927.
[34] S. Liu, D. Yi, Z. Lei, and S. Z. Li, "Heterogeneous face image matching using multi-scale features," in *Fifth IAPR International Conference on Biometrics*. IEEE, 2012, pp. 79–84.
[35] J. Li, K. Lu, Z. Huang, L. Zhu, and H. T. Shen, "Transfer independently together: A generalized framework for domain adaptation," *IEEE Transactions on Cybernetics*, 2018.
[36] H. Liu, M. Shao, Z. Ding, and Y. Fu, "Structure-preserved unsupervised domain adaptation," *IEEE Transactions on Knowledge and Data Engineering*, 2018.
[37] J. Li, Y. Wu, J. Zhao, and K. Lu, "Low-rank discriminant embedding for multiview learning," *IEEE transactions on cybernetics*, vol. 47, no. 11, pp. 3516–3529, 2017.
[38] M. R. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.
[39] X. Xu, W. Li, D. Xu, and I. W. Tsang, "Co-labeling for multi-view weakly labeled learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 6, pp. 1113–1125, 2016.
[40] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.
[41] Y. Wen, Z. Li, and Y. Qiao, "Latent factor guided convolutional neural networks for age-invariant face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4893–4901.
[42] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
[43] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems*, 2014, pp. 3320–3328.
[44] Z. Ding, S. Suh, J.-J. Han, C. Choi, and Y. Fu, "Discriminative low-rank metric learning for face recognition," in *12th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*. IEEE, 2015.
[45] F. Nie, S. Xiang, Y. Jia, C. Zhang, and S. Yan, "Trace ratio criterion for feature selection." in *Association for the Advancement of Artificial Intelligence*, 2008, pp. 671–676.
[46] Y. Jia, F. Nie, and C. Zhang, "Trace ratio problem revisited," *IEEE Transactions on Neural Networks*, vol. 20, no. 4, pp. 729–735, 2009.
[47] Z. Kang and Q. Cheng, "Top-n recommendation with novel rank approximation," in *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 2016, pp. 126–134.
[48] L. Mackey, A. Talwalkar, and M. I. Jordan, "Distributed matrix completion and robust factorization," *Journal of Machine Learning Research*, vol. 16, pp. 913–960, 2015.
[49] Y. Pan, R. Xia, J. Yin, and N. Liu, "A divide-and-conquer method for scalable robust multitask learning," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 12, pp. 3163–3175, 2015.
[50] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," in *Proceedings of the Nineteenth annual ACM Symposium on Theory of Computing*. ACM, 1987, pp. 1–6.
[51] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *IEEE International Conference on Computer Vision*, 2011, pp. 1615–1622.
[52] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
[53] R. Horst and N. V. Thoai, "Dc programming: overview," *Journal of Optimization Theory and Applications*, vol. 103, no. 1, pp. 1–43, 1999.



Zhengming Ding (S'14) received the B.Eng. degree in information security and the M.Eng. degree in computer software and theory from University of Electronic Science and Technology of China (UESTC), China, in 2010 and 2013, respectively. He received the Ph.D. degree from the Department of Electrical and Computer Engineering, Northeastern University, USA in 2018. He is a faculty member affiliated with Department of Computer, Information and Technology Indiana University-Purdue University Indianapolis since 2018. His research interests include machine learning and computer vision. Specifically, he devotes himself to develop scalable algorithms for challenging problems in transfer learning and deep learning scenario. He received the National Institute of Justice Fellowship during 2016-2018. He was the recipients of the best paper award (SPIE 2016) and best paper candidate (ACM MM 2017).



Yun Fu (S'07-M'08-SM'11) received the B.Eng. degree in information engineering and the M.Eng. degree in pattern recognition and intelligence systems from Xi'an Jiaotong University, China, respectively, and the M.S. degree in statistics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign, respectively. He is an interdisciplinary faculty member affiliated with College of Engineering and the College of Computer and Information Science at Northeastern University since 2012. His research interests are Machine Learning, Computational Intelligence, Big Data Mining, Computer Vision, Pattern Recognition, and Cyber-Physical Systems. He has extensive publications in leading journals, books/book chapters and international conferences/workshops. He serves as associate editor, chairs, PC member and reviewer of many top journals and international conferences/workshops. He received seven Prestigious Young Investigator or Early Career Awards from NAE, ONR, ARO, IEEE, INNS, UIUC, Grainger Foundation; nine Best Paper Awards from IEEE, IAPR, SPIE, SIAM; four major Industrial Research Awards from Google, Samsung, Mathworks and Adobe, etc. He is currently an Associate Editor of the IEEE Transactions on Neural Networks and Leaning Systems (TNNLS) and IEEE Transactions on Image Processing (TIP). He is fellow of IAPR, fellow of SPIE, a Lifetime Senior Member of ACM, Lifetime Member of AAAI, OSA, and Institute of Mathematical Statistics, member of ACM Future of Computing Academy, Global Young Academy (GYA), INNS and Beckman Graduate Fellow during 2007-2008.